Phenotypic Characterization of Microorganisms by Barcoded Transposon Mutagenesis

Adam Deutschbauer[1,4] (AMDeutschbauer@lbl.gov), Julia Oh[2], Morgan Price[1,4], Paramvir Dehal[1,4], Dan Bruno[2], Jennifer Kuehl[1,4], Romy Chakraborty[1,4], Terry C. Hazen[1,4], Corey Nislow[3], Guri Giaever[3], Ronald W. Davis[2], Adam P. Arkin[1,4]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Stanford Genome Technology Center, Stanford, CA; [3]University of Toronto, Toronto, ON, Canada; [4]VIMSS (Virtual Institute of Microbial Stress and Survival) http://vimss.lbl.gov/

ESPP2 Environmental Stress Pathway Project

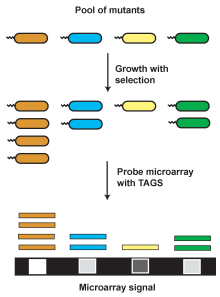
DOE GENOMICS:GTL ACCELERATING DISCOVERY FOR ENERGY AND ENVIRONMENT

VIMSS Virtual Institute for Microbial Stress and Survival | Massachusetts Institute of Technology | University of Missouri | University of Washington | Montana State University | Southern Illinois University Carbondale | University of Oklahoma | Sandia National Laboratories | Oak Ridge National Laboratory | Berkeley University of California | Lawrence Berkeley National Laboratory | http://vimss.lbl.gov/

Introduction

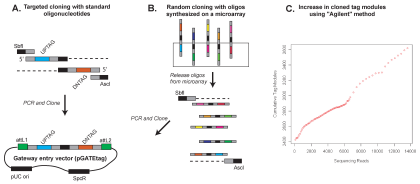
Systems-level analyses of less studied bacteria are limited by the presence of numerous uncharacterized genes and an over reliance on annotations from well studied bacteria such as *E. coli*. To meet this challenge, we are developing a mutagenesis and phenotyping strategy that is comprehensive across the genome and applicable to any microorganism amenable to transposon mutagenesis. We have cloned and sequence-verified ~3600 barcode modules into a Gateway entry vector. Each module is a 175 base pair element containing two unique 20 base pair sequences, the UPTAG and DOWNTAG, flanked by common PCR priming sites. Each module can then be rapidly transferred *in vitro* to any DNA element, such as a transposon, that is made Gateway compatible. Transposon mutants marked by the modules will be sequenced to determine which of the ~3600 barcode modules was used and which gene was disrupted. Transposon mutants can be rapidly re-arrayed into a single pool containing ~3600 uniquely tagged, sequence-verified mutant strains. By sequencing saturating numbers of transposon mutants, we can identify and assay mutants in most nonessential genes in a given genome. The fitness of each mutant in the pool will be monitored by the hybridization of the barcodes to an Affymetrix microarray containing the tag complements in a system identical to that used for the yeast deletion collection. Compared to other approaches for the parallel anlaysis of transposon mutants such as signature tagged mutagenesis, genetic footprinting, and transposon site hybridization, our approach offers higher throughput, a single microarray design is universal for any organism, single mutational events are assayed, and mutant strains are archived for verification, further analysis, and distribution. Here we describe the application of this approach to the genomes of *Shewanella oneidensis* MR1 and *Desulfovibrio desulfuricans* G20. The completion of this project will enable the quantitative phenotypic analysis of thousands of mutants across a wide range of conditions. In addition, our genetic resources provide a framework for the systematic genetic interrogation of individual pathways.

Parallel Analysis of Mutant Pools



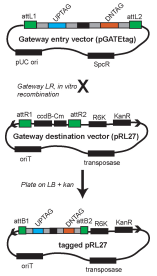
A pool is constructed by combining a collection of mutants, each carrying a unique UPTAG and DOWNTAG. The relative abundance of all strains in the pool can be monitored by the hybridization of the UPTAGs and DOWNTAGs to an Affymetrix microarray containing the tag complements. This system has been successfully used to monitor the fitness of ~6000 yeast deletion strains across thousands of conditions. We aim to make this strategy universal for virtually any genetically amenable microorganism.

A universal tag module collection



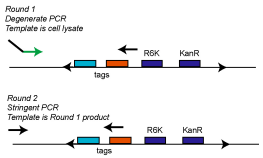
(A) The tag modules are constructed using two long oligonucleotides, one oligo contains the UPTAG, a second oligo contains the DOWNTAG. Each tag module is cloned into a modified Gateway entry vector and sequence-verified. 2400 tag modules were made using this method.
(B) ~6000 oligos not cloned by method A were synthesized on a microarray (Agilent) and cloned randomly. The tag module identity and accuracy were verified by sequencing.
(C) Increase in new tag modules by "Agilent" method. We aim to make ~5000 tag modules.

Tag transfer via Gateway reaction



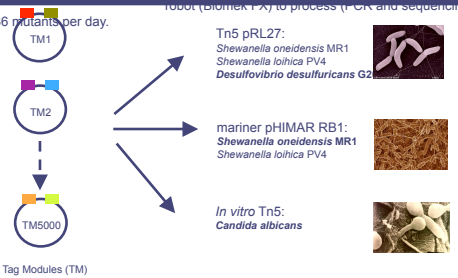
The tag modules are universal. They can be transferred *in vitro* to any Gateway compatible destination vector (either as single reactions or in pools). We converted the Tn5-transposon delivery plasmid pRL27 into a Gateway entry vector by cloning the "AttR1-ccdB-CmR-AttR2" cassette into a KpnI site located at the end of Tn5.

Transposon insertion mapping



- After random mutagenesis with the collection of tagged transposons, we determine the insertion location of the transposon in the genome and the tag identify using a 2-step degenerate PCR approach. Sequence analysis of the second round PCR product is used to infer both the tag identify and the genome insertion site.
- We anticipate sequencing ~30,000 colonies per bacterial genome in order to get adequate coverage. Mutants with unique insertion sites will be selected from these libraries and combined into pools of ~5000 mutants (each with a unique tag module) robot (Biomek FX) to process (PCR and sequencing) 1536 mutants per day.

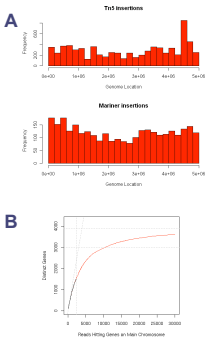
Our focus organisms



Tag Modules (TM)

We have converted pRL27 (Tn5), pHIMAR-RB1 (mariner), and an *in vitro* Tn5 system (epicentre) into Gateway destination vectors. The tagged transposons are functional in the indicated organisms.

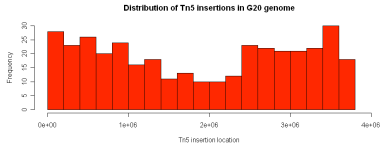
Application to *Shewanella oneidensis* MR1



(A) Tagged transposon insertion distribution for 5762 Tn5 mutants and 2297 mariner mutants on the main chromosome of *Shewanella oneidensis* MR1. Both transposons have a significant insertion biases. Currently, we have mapped insertions into 2740 unique protein-coding genes in the *Shewanella* MR1 genome. Assuming that 10% of the genome is essential for viability, we have targeted ~70% of the nonessential *Shewanella* MR1 genome. The 2740 unique genes include 2003 distinct tag modules. Currently, our ability to make a single deletion pool is tag (not gene) limited.

(B) Simulation of how many sequencing reads it would take to hit the entire nonessential genome (assumed to be ~3900 genes). The simulation only takes into account the observed insertion distribution for the mariner transposon. The black line represents the observed data; the red line represents the simulation out to 30,000 useful reads.

Application to *Desulfovibrio desulfuricans* G20



- In a pilot study, 432 tagged Tn5 mutants were sequenced. We determined the tag identify and insertion location for 368 mutants (85% of the attempted reads). Shown is the insertion location for these mutants on the *Desulfovibrio desulfuricans* G20 genome. There is a bias towards insertions near the origin of replication and away from the terminus

Future directions

- Generate ~1000 additional sequence-verified tag modules.
- Go into production phase and rapidly generate thousands of sequence-defined transposon mutants in multiple microbial genomes.
- Explore the potential for double deletion construction for global genetic interaction studies.
- Automate and miniaturize mutant pool experiments using procedures developed for the yeast deletion collection. Such an effort will enable us to profile comprehensive mutant libraries across hundreds of diverse conditions.

ACKNOWLEDGEMENT

ESPP2 (MDCASE) is part of the Virtual Institute for Microbial Stress and Survival (VIMSS) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL Program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.